University of Hertfordshire UH

# Clustering and Data Fitting

**Name:** Anush Farooq

**Student ID:** 22026480

Github Repository Link: https://github.com/Anush123678/assignment_3_clustering.git

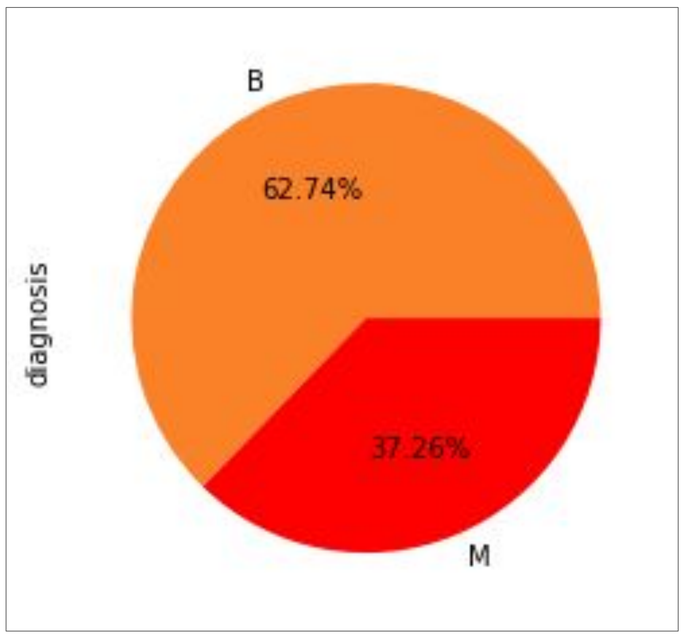
**Abstract**

Cancer is the most dangerous and incurable disease (if the cancer reached at its peak). To classify the types of cancer such that the cancer is benign or malignant, I have studied some of the properties of cancer. I have made comparison for different characteristics of human cell nuclei to classify the type of cancer.

# Classification of Benign vs. Malignant cancer and characteristics of cell nuclei

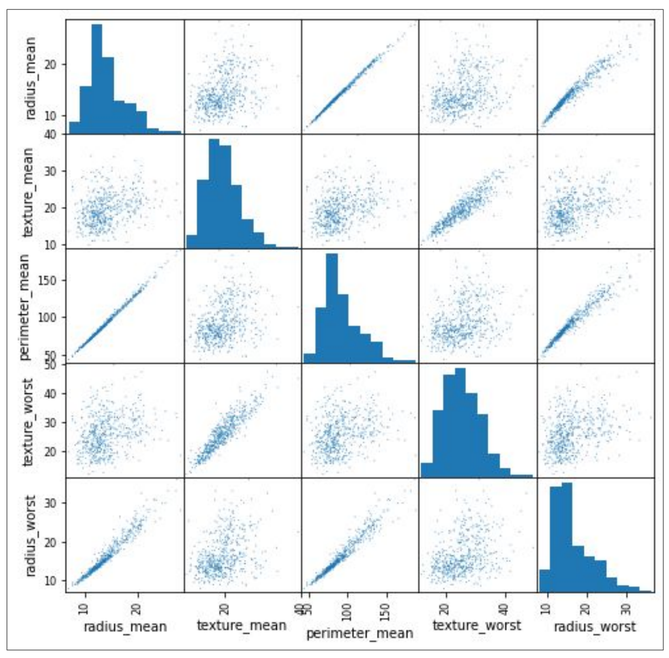
## Summary


- In these days, there is a great need to think about cancer and its main causes as it is a severe cause of death. There is no medicine till now for the recovery of any kind of cancer.
- The presence of cancer in the human body can be predicted by observing the human cell nuclei and change in these cells.
- A sample of human blood/mass cell can be obtained by the process of fine needle aspirate (FNA)[1]. In order to collect this sample, a needle is inserted into the human mass.
- The cancer disease is much common nowadays than ever due to the following reasons:[2]
  - Use of tobacco
  - Alcohol consumption
  - Air pollution
  - Unhealthy diet

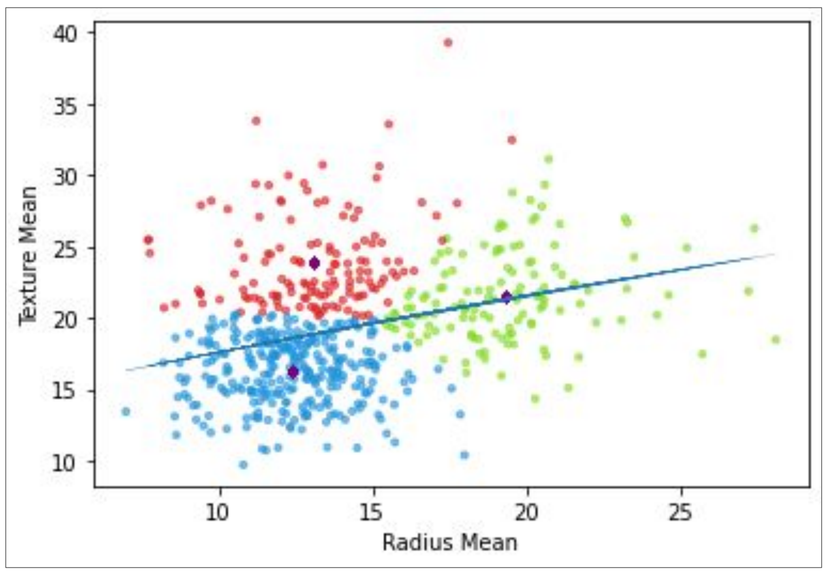

The preceding pie plot shows the distribution of benign and malignant cancer. We can observe that there is not a big difference in the ratio of both levels of cancer.

## Introduction




- In the above plot (scatter matrix), we can observe the most of the sell characteristics are dependent on each other. If any of the property got changed, it will affect the complete cell of human body.
- To precede our study against cancer and the characteristics of the affected human cell nuclei, we have chosen main two parameters to dig into them. The radius of cell and texture of cell has great impact in the causing of cancer disease since we can also confirm it from the above scatter matrix plot.

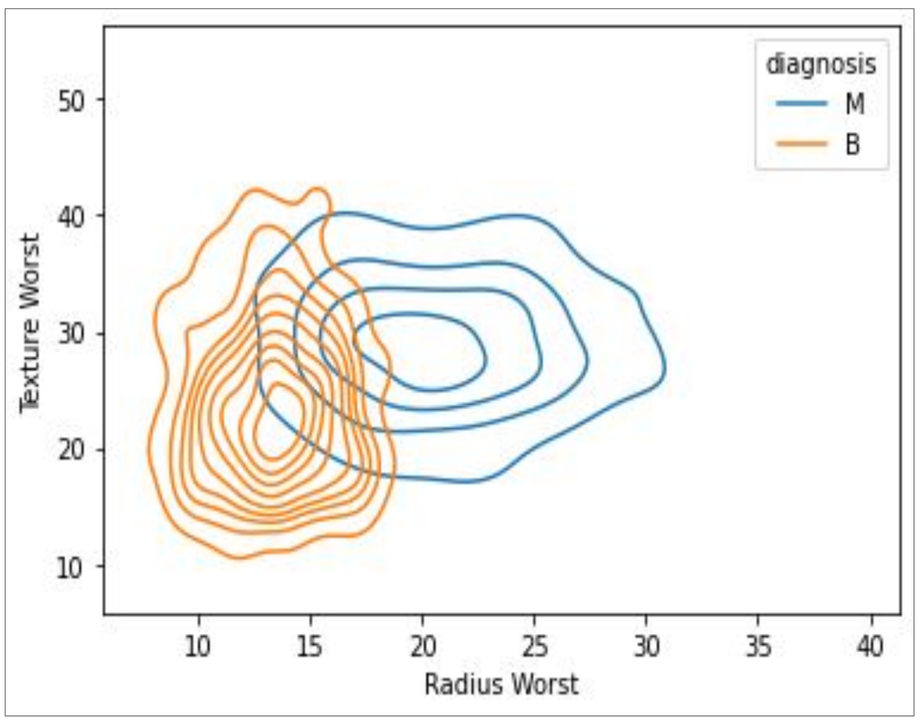
## Cell nuclei Classification




- To proceed my research, I classified the human cell nuclei in three clusters.
- After applying the curve fit on this data, I observed that this data shows some kind of linear relationship between mean of cell nuclei radius and mean of cell nuclei texture.

## Disease Diagnosis in Classification

- The KDE (kernel density estimate) plots are used to visualize the distribution of observations in a dataset, comparable to a histogram.
- In the following KDE plot, we can observe an intersection of benign and malignant cancer levels.
- The human who is suffering from the cancer disease in the benign level, can have much chances that this cancer can be  a malignant if the cell nuclei has worst radius between 13 to 20 units and texture worst between 18 to 50 units.




In this image, we can observe the overlapping area of the classification of cancer levels. The human suffering from cancer and having the cells characteristics in this area should take some precautions so that the cancer will not change its level to malignant.

## References


- [1] https://www.enthealth.org/conditions/fine-needle-aspiration/
- [2] https://www.who.int/news-room/fact-sheets/detail/cancer


University of Hertfordshire UH

THE WORLD BANK
IBRD • IDA